

Experimental Adaptive Bayesian Tomography

K.S.Kravtsov¹, S.S.Straupe^{2,*} and I.V.Radchenko¹, N.M.T.Houlsby³, S.P.Kulik²

¹ *A.M.Prokhorov General Physics Institute RAS, Moscow, Russia*

² *Faculty of Physics, M.V.Lomonosov Moscow State University, Moscow, Russia*

³ *Computational and Biological Learning Lab, Department of Engineering,
University of Cambridge, Cambridge, United Kingdom*

(Dated: March 8, 2013)

We report an experimental realization of an adaptive quantum state tomography protocol. Our method takes advantage of a Bayesian approach to statistical inference and is naturally tailored for adaptive strategies. For pure states we observe close to $1/N$ scaling of infidelity with overall number of registered events, while best non-adaptive protocols allow for $1/\sqrt{N}$ scaling only. Experiments are performed for polarization qubits, but the approach is readily adapted to any dimension.

PACS numbers:

The main goal of quantum state tomography is to provide an estimate $\hat{\rho}$ for an unknown quantum state ρ based on the data collected in a series of measurements [1]. The estimator is supposed to be close to the real state in some reasonable sense, therefore various notions of statistical distance between quantum states are used [2, 3]. One of the possible measures of statistical distance is *infidelity* [22], defined as $1 - F(\rho, \hat{\rho}) = 1 - \text{Tr}(\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}})^2$. The ultimate goal of any tomographic protocol is to minimize this distance for a fixed overall number of measurements made N . Usually a protocol makes use of some fixed number of measurement settings predetermined before the actual experiment. For such a protocol the infidelity scales as $1 - F \sim N^{-1/2}$ for the most interesting for applications set of almost-pure states. One can more or less significantly alter the pre-factor by a clever choice of measurements [5–7], but the scaling is unaffected. A natural question is whether it is possible to beat this limit? The answer turns out to be positive if one allows for adaptivity - the measurement performed at some step of the protocol should be determined in dependence of the data obtained in the previous ones [8, 9].

Here we report an experimental approach to adaptive quantum state tomography based on a recently proposed adaptive Bayesian estimation algorithm [10]. We achieve almost $1/N$ scaling of infidelity for pure states of polarization qubits and demonstrate a clear advantage over best symmetric non-adaptive protocols. Our approach is completely different and more general than that of another recent experimental realization [11], where adaptive measurements were used to estimate a single unknown parameter of a quantum state.

Bayesian tomography. Let us start with describing a general framework for quantum state estimation and, in particular, the Bayesian approach. A tomographic protocol is a set of positive operator valued measures (POVM's) $\mathcal{M} = \{\mathbb{M}_\alpha\}$ with index α numbering the different configurations of the experimental apparatus. In a given configuration, the probabilistic outcome of each measurement γ being observed is determined according

to the Born's rule:

$$\mathbb{P}(\gamma|\rho, \alpha) = \text{Tr}[M_{\alpha\gamma}\rho], \quad (1)$$

where $M_{\alpha\gamma}$ are POVM elements, obeying $\sum_{\gamma=0}^{\Gamma-1} M_{\alpha\gamma} = I$, and ρ is the density matrix of the state to be determined. The set \mathcal{D} of all outcomes observed in an experiment form the data set used to estimate density matrix elements. The Bayesian approach to statistical inference dictates the following rules:

- a *prior* distribution over the space of density matrices $p(\rho)$ is specified;
- the collected data are used to obtain the posterior distribution $p(\rho|\mathcal{D}) \propto \mathcal{L}(\rho; \mathcal{D})p(\rho)$, $\mathcal{L}(\rho; \mathcal{D})$ is the *likelihood* function, and it contains our statistical model that encodes probabilistic mapping from the state to the observed data;
- quantities of interest are estimated using expected values under the posterior distribution: for example, we may obtain the *Bayesian mean estimate* of the state as $\hat{\rho} = \mathbb{E}_{p(\rho|\mathcal{D})}[\rho]$. Variance, infidelity or any other statistical quantity of interest may be obtained in the same way.

The Bayesian approach has many advantages over a more standard for quantum information community maximum-likelihood estimation (MLE) [4]. It offers, in a natural way, a distribution over the space of density matrices, which provides the most complete description of our knowledge about the quantum state, inferred from data \mathcal{D} [12]. Even more importantly for us, it is a natural framework for construction of adaptive estimation protocols. Indeed, the posterior distribution may be updated as soon as one observes some data, in the extreme case - after each measurement, and the new knowledge about the state may be used to select the next measurement setting α in a most optimal way. Choosing the criterion for "optimality" is a task of *optimal experiment design* (OED) and may be solved in various ways. In Bayesian

framework a natural strategy is to choose a measurement, maximally reducing the entropy of the posterior - it means, that our knowledge about the state, obtained after such measurement is maximized [10]. This may be formulated as choosing a measurement configuration α as a solution to the following optimization procedure:

$$\alpha = \arg \max_{\alpha \in \mathcal{A}} \{ \mathbb{H}[p(\rho|\mathcal{D})] - \mathbb{E}_{p(\gamma|\alpha, \mathcal{D})} \mathbb{H}[p(\rho|\gamma, \alpha, \mathcal{D})] \}. \quad (2)$$

Note, that because we do not know which outcome γ will be observed we use the *expected* information gain (under the posterior) as our objective. Before describing how to work with (2) in practice we detail the components required for our Bayesian model.

The likelihood function. The likelihood function is equal to the probability of the observed data, given a particular state, i.e. $\mathcal{L}(\rho; \mathcal{D}) = \mathbb{P}(\gamma|\rho, \alpha)$. In the simplest setting, i.e. in the absence of any experimental noise, the likelihood function is given directly by Born's rule (1). In practice experimental noise also needs to be modelled in the likelihood function, we present these extensions to the simple model later in the paper.

The prior. As the Bayesian framework implies finding a probability distribution instead of a single point estimate, the analysis should also take into account the particular geometry of the space, i.e. the geometry of single qubit density matrix space. In general, the geometry of space is defined by its metric, which provides a notion of distance. In the case of density matrices the natural choice of metric is Bures distance, defined as $d_B^2(\rho_1, \rho_2) = 2 - 2\sqrt{F(\rho_1, \rho_2)}$. Locally it coincides with the concept of Fubini-Study (quantum angle) measure and the Hilbert-Schmidt distance (see Ref. [3]). As follows from its definition, for close-by states the Bures distance is a square root of infidelity.

It can be shown that the curvature of Bures metric space for single qubit states is constant [13], thus it can be isometric to a hypersphere in a 4-d space. A simple isometric mapping exist between the hemisphere of radius 1/2 and the Stokes parameters:

$$\begin{aligned} x_1 &= \frac{1}{2}S_1; & x_2 &= \frac{1}{2}S_2; & x_3 &= \frac{1}{2}S_3; \\ x_4 &= \frac{1}{2}\sqrt{1 - S_1^2 - S_2^2 - S_3^2}. \end{aligned} \quad (3)$$

Thus a uniform distribution in the Bures distance sense is a projection of a uniformly populated 3-sphere to the space of Stokes parameters. This projection apparently gives a lower density in its center and a higher near its surface.

The first step in any Bayesian estimation procedure is to choose a prior distribution. In the ideal situation a prior should give absolutely no information about the system, thus a typical choice for prior is a Haar measure in the space in question. In our case we use Haar measure in Bures metric space. However, to stick with conventional (and very convenient) parameterization with three

Stokes parameters we use the above mentioned isometry from a 3-sphere. Samples obtained by this procedure concentrate more along the surface of the Stokes parameters ball, which is natural in the sense that the distance and, thus, *fidelity* between the samples remains uniform over the whole space. This would not be the case if the ball was populated uniformly, as fidelity is not directly connected with separations in the space of Stokes parameters.

We should also mention here another strategy extensively used in literature for prior generation [12]. According to it a density matrix for a single-qubit is treated as a result of tracing out other qubits from a higher-dimensional pure state. Starting from different dimensions one gets different prior distributions, however with the increase of dimensionality the distribution tends to concentrate around the completely mixed state giving less and less chances to pure states. This unnatural behavior prevents us from using it.

To illustrate the above, in Fig. 1 we show density of samples in a flat disc cut from the center of the Stokes parameters ball for the Haar measure (a), and for the results of tracing down from a two-qubit (b) and a qubit-qutrit (c) pure states. The first distribution has more density towards the circumference following the behavior of fidelity in the Stokes parameter space. Other distributions give more and more favor to mixed states, unveiling their inadequateness for our purpose.

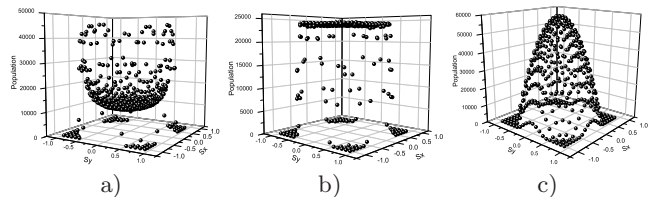


Figure 1: Population of samples in the space of Stokes parameters vs. S_x and S_y condition to $|S_z| < 0.05$. Samples are derived from (a) Haar measure in Bures metric space; two qubit (b) and qubit-qutrit (c) pure states traced by the second particle. The Haar measure gives flat infidelity distribution between all samples, while the latter two favor mixed states giving poorer fidelity to pure states.

Approximate Inference One of the principle reasons that Bayesian methods enjoy less popularity in quantum tomography than MLE is the fact that posterior normalization requires computing an (in general high-dimensional) integral of likelihood function, which is computationally hard. Usually, when faced with intractable Bayesian inference, the posterior is approximated via sampling [16], or by approximating the posterior with a simpler distribution, such as a Gaussian [17].

This computational difficulty is further compounded when one is performing adaptive quantum tomography, one must keep track of the current posterior after making

n measurements, in order to calculate the optimal $n+1$ 'st measurement. To perform inference based on all the observed data is at best an $\mathcal{O}(n)$ operation, which becomes increasingly problematic as the experiment progresses. Fortunately fast algorithms for solving online Bayesian inference problems exist; they update the posterior after inclusion of each new datapoint, without re-visiting all previous data. We briefly review the core idea behind this approach, and refer the reader to [18] for details.

The algorithm is a variant of sequential importance sampling algorithm (SIS) with resampling. The idea is to construct a particle filter, approximating the posterior with a set of weighted samples, i.e. $p(\rho|\mathcal{D}_n) \approx \sum_{s=1}^S w_s^{(n)} \delta(\rho - \rho_s)$. After each observation one updates the weights $w_s^{(n)}$; this can be done incrementally, using the current set of particles and weights, and the likelihood corresponding to the new observation. The algorithm has $\mathcal{O}(1)$ cost for integration of the likelihood, which means that it can be applied on-line at every step of the adaptive protocol, irrespective of the current amount of data collected.

A common problem with weighted particle filters is that after collection of sufficient data, a small collection of particles get almost all of the weight, this means that the *effective sample size* reduces and the quality of the approximation to the posterior becomes poor. This problem is avoided by monitoring the effective sample size and re-sampling the particles when it falls too low. The resampling procedure uses two phases, firstly the particles are re-drawn from the set of current particles in proportion to their weight (and the weights are re-set to zero), then they are ‘spread back out’ using the Metropolis-Hastings algorithm. It is important to note that implementation of Metropolis-Hastings algorithm is convenient to perform on a 3-sphere surface defined by (3), which automatically ensures correct step sizes and avoids unnecessary boundary conditions on the surface of the Bloch sphere, which, in general, may bias the posterior. Using the particles, whenever we require an expectation under the posterior, e.g. when computing the mean fidelity, or when computing the next state for adaptive tomography, one can simply replace the complex integrals with simple weighted Monte Carlo estimates.

Efficient Adaptive Tomography We now return to computation of the objective function for adaptive tomography (2), although this objective is theoretically attractive, even with the sampling estimate of the posterior two major computational difficulties are encountered. Firstly, we must compute entropies of (in general) high-dimensional quantum states; it is notoriously hard to compute entropies directly from samples from the distribution [19]. Secondly, one requires the posterior distribution $p(\rho|\gamma, \alpha, \mathcal{D})$ for all possible next measurements γ , and all their possible outcomes α ; this would require performing an SIS update for all these possible scenarios.

This computational burden is unavoidable if one uses a loss function other than the log loss, which leads to the Shannon’s entropy objective function, which means that optimal designs can only be computed for very short experiments [20]. Therefore, it is highly beneficial to work with the following equivalent formulation:

$$\alpha = \arg \max_{\alpha \in \mathcal{A}} \{ \mathbb{H}[p(\gamma|\alpha, \mathcal{D})] - \mathbb{E}_{p(\rho|\mathcal{D})} \mathbb{H}[p(\gamma|\alpha, \rho)] \}. \quad (4)$$

In (4) only predictive entropies are required, which is much easier because output space is typically much lower dimensional than state space, and only the *current* posterior is needed $p(\rho|\mathcal{D})$.

Simulations We performed simulated experiments to empirically evaluate the performance of Bayesian adaptive tomography. For our performance metric we use the mean infidelity as measured against the true state, $\bar{\rho}$: $1 - \hat{F}(\rho, \bar{\rho}) = \mathbb{E}_{p(\rho|\mathcal{D}_n)} [1 - F(\rho, \bar{\rho})]$. Note that Bayesian mean $1 - \hat{F}(\rho, \bar{\rho})$ is a ‘fairer’ score than the fidelity of a point estimate, e.g. the posterior mean (i.e. $[1 - F(\mathbb{E}_{p(\rho|\mathcal{D}_n)}[\rho], \bar{\rho})]$). The fidelity of the posterior mean does not take into account the uncertainty captured by the posterior. The posterior mean could be correct, for example, if the state is pure and the posterior has become ‘flattened’ against the surface of the Bloch sphere; the variance, however, could be very high - the system may have little knowledge of the polar and azimuth angles (in the Bloch sphere) of the true state. The Bayesian estimator rewards posterior distributions that are centered in the correct location *and* have low variance.

To achieve statistically significant results we perform multiple runs within each simulation. For each run we generate a random pure state which we use as the true state $\bar{\rho}$. We average over 20 runs, each with a different true state. All measurements performed in a single run are shown in Fig. 2. After the system collects some initial information about the measured state it tends to choose measurements aligned with its current estimate of the true state (although not exclusively), thus taking advantage of the adaptive approach. Interestingly, the optimal first three measurements chosen correspond precisely to a set of Mutually Unbiased Bases (MUBs), but diverge from these MUBs on the 4th. We compare the results to measurements chosen randomly and uniformly, and selecting randomly from a set of MUBs, which have been shown to be the optimal (in terms of information gain) non-adaptive measurement set one can choose prior to the experiment [21]. We also fit a power law, of the form $1 - \hat{F} \propto N^a$ to the data, in order to compare the convergence rates of the different methods. Note that the performance of the random and adaptive schemes is independent of the angle of the true state (the algorithms remain the same given any rotation of the Stokes coordinate), however, the fixed MUBs are not. Drawing intuition from the fact that the adaptive scheme selects mostly mea-

measurements that align with the true state (hence ‘squashing’ the posterior against the Bloch sphere), we find that the ‘best case’ for MUB tomography is when the state is aligned with one of the MUB measurements, and the worst case is when the state is equally biased to all measurements i.e. $\{\sigma_x = \sigma_y = \sigma_z = 1/\sqrt{3}\}$.

The results are presented in Fig. 3. Firstly, note that random tomography yields a $1/\sqrt{N}$ rate as expected, $\alpha = 0.448 \pm 0.183$. However, adaptive tomography performs close to the $1/N$ level on average $\alpha = 0.915 \pm 0.101$. In its most favorable scenario, MUBs also perform close to the $1/N$ rate (with a small multiplicative constant improvement over the adaptive scheme). In practice, because the optimal MUBs depend on the state, these are not known a priori. In the worst case MUBs achieve $1/\sqrt{N}$ convergence; but we also observe that on average the rate is closer to $1/\sqrt{N}$ than $1/N$, $\alpha = -0.593 \pm 0.134$.

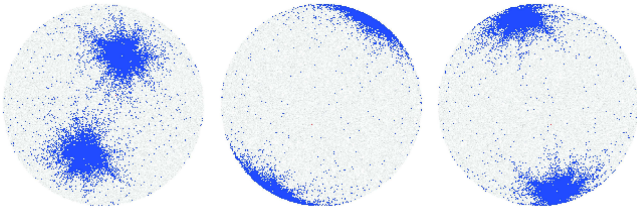


Figure 2: Illustration of an adaptive choice of measurements at reconstruction of a pure state. The Bloch sphere is given in three orthogonal projections with each measurement basis marked as a dark point. As expected, measurement bases tend to concentrate around the reconstructed state or the symmetric state on the other side of the sphere.

Measurement blocks. One can save computational and experimental time by taking blocks of measurements using the same configuration for k consecutive measurements. As the posterior collapses towards the true state, the optimal measurement changes less frequently - a direction pointing towards the true state becomes increasingly preferable. Therefore, as the experiment progresses the block size k can be allowed to grow without detrimenting the quality of the experimental procedure. We use a heuristic block-size schedule which increases the block size at a $\mathcal{O}(n)$ rate, where n is the number of measurements seen so far. In particular we use $k = \max(\lfloor n/100 \rfloor, 1)$; in this case the achieved infidelity scales linearly with elapsed time. Simulations show that using this schedule does not make any statistically significant difference to the convergence rate, or even to the absolute fidelity achieved at any time.

Experimental imperfections. In practice quantum tomography is inevitably subject to experimental noise. This noise is not modeled in the likelihood function given by Born’s rule (1), and therefore, its presence may bias the results of inference, reducing both the fidelity of the inferred state, and the optimality of the adaptive experi-

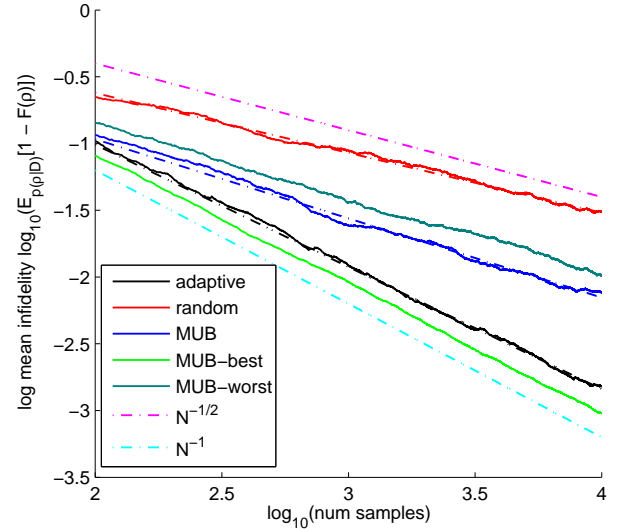


Figure 3: Simulated tomography using three measurement selection methods, randomly sampled (red), MUBs (blue) and fully adaptive Bayesian tomography (black). For these methods, the true state is random and pure, the results presented here are the average of 20 independent runs. Overlaid dashed lines indicate the power law fit. Functions $1 - F = cN^{-1/2}$ (magenta) and $1 - F = cN^{-1}$ (cyan) are shown for comparison. The performance of random and adaptive tomography is clearly independent of the Bloch-sphere angle of the pure state, the fixed MUBs are not. Therefore, we also present the performance of MUB tomography given the ‘worst’ and ‘best’ (see text) true states (dark, light green respectively).

mental design. For our set-up we have identified two major additional sources of experimental noise. Firstly, the presence of detector dark counts with detector-specific rates. Secondly, attenuation in both channels due to detector inefficiency and losses/reflections at the optical elements. If the attenuation was equal in both channels, then the inference would be unaffected; however, unequal attenuation will bias the posterior.

A popular approach to modeling the additional uncertainty in the state is to model the observed state as a linear mixture of the true state and the maximally mixed state [15]. Although with this assumption one can model certain simple noise sources, such as equal dark counts arriving with equal rates at each detector, we address the specific sources of noise in our experimental paradigm more directly. To model dark counts, we assume that the production of photons by the laser source, and the arrival of dark counts at the detectors can be modeled using independent homogeneous Poisson processes with (constant) rate parameters λ_s for the source and λ_d^j for each detector. These rates are estimated a priori using a pilot experiment. From these assumptions one can derive the following likelihood function using the properties of the

Poisson distribution:

$$\mathbb{P}(\gamma|\rho, \alpha, \lambda_s, \lambda_d^1, \dots, \lambda_d^\gamma) = \frac{\text{Tr}[M_{\alpha\gamma}\rho]\lambda_s + \lambda_d^\gamma}{\lambda_s + \sum_\gamma \lambda_d^\gamma} \quad (5)$$

To deal with channel efficiency, we assume that there is a fixed probability of a photon being ‘lost’ from a channel; this probability is denoted $1 - \eta_\gamma$ for each channel γ , and hence η_γ may be interpreted as the ‘channel efficiency’. These probabilities are also estimated in a preliminary experiment. Given these efficiencies the likelihood becomes:

$$\mathbb{P}(\gamma|\rho, \alpha, \eta_1, \dots, \eta_\gamma) = \frac{\text{Tr}[M_{\alpha\gamma}\rho]\eta_\gamma}{\sum_\gamma \text{Tr}[M_{\alpha\gamma}\rho]\eta_\gamma} \quad (6)$$

Note that in both cases, both the numerator and denominator contain only linear terms in the additional parameters (λ, η) . Therefore, one only requires estimates of the ratio of the dark count rates to the source rate $\lambda_d^\gamma/\lambda_s$, and, for single-qubit tomography, the ratio of the efficiencies of the two channels η_1/η_2 . It is straightforward to show that this property also holds when one collects blocks of measurements in one configuration and computes the likelihood of all measurements in the block simultaneously.

Experiment. A sketch of our experimental setup is shown in Fig. 4. We use a CW 850 nm VCSEL diode laser coupled to a single-mode fiber as a light source. The radiation is attenuated to the single-photon level by a set of neutral density filters F and additionally spatially filtered with small iris apertures. The input polarization state is defined by a Glan-Taylor prism GP with high extinction ratio (more than 6000:1), the prism transmits horizontally polarized light, which may be transformed to some arbitrary state with a custom quartz plate WP.

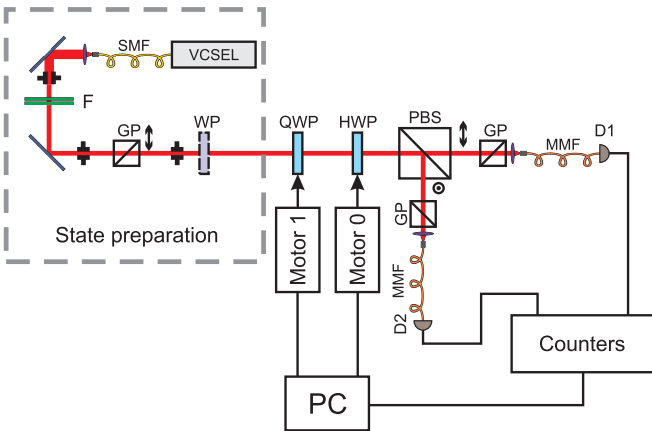


Figure 4: Experimental setup. An attenuated laser is used as a source, the polarization state is prepared by a set of waveplates, and analyzed by a sequence of a quarter- and half-wave plates, followed by a polarizing beam-splitter and two single-photon counters. Waveplates are rotated by electronically controlled step-motor drivers to allow for adaptivity.

The measurement scheme consists of zero order quarter-wave plate QWP and half-wave plate HWP. The plates are rotated by step-motor-driven stages, with minimal angular step of 0.1° . The zero position is controlled by a Hall sensor providing uncertainty of wave-plates zero of 0.2° . We clean up the polarization states in the output channels of a polarization beam-splitting cube (PBS) with two additional Glan-Taylor prisms to ensure high extinction ratio. Effectively that is equivalent to introducing some losses in the non-ideal PBS without altering the output polarization states. In each channel photons are coupled to multi-mode fibers and detected by single photon counting modules D1 and D2 (Perkin-Elmer). Electronic pulses from SPCM’s are sent to home-made counters which may operate in two regimes - count for a fixed period of time or count until the specified number of counts is reached.

To show the advantage of adaptive state estimation over non-adaptive protocols we performed a direct comparison in the Bayesian framework. All non-adaptive protocols were shown to scale similarly in the limit of large N in our simulations, so we chose randomly sampled measurements for comparison. In the adaptive estimation scheme we used two strategies: adaptation after every single measurement and block measurements and found no statistically significant differences. Fig. 5 shows the dependence of mean infidelity $1 - \hat{F}(\rho, \hat{\rho}) = \mathbb{E}_{p(\rho|\mathcal{D}_n)}[1 - F(\rho, \hat{\rho})]$ with current estimate $\hat{\rho}$ (for which we used the Bayesian posterior mean) on the number of measurements performed. It is important to note that

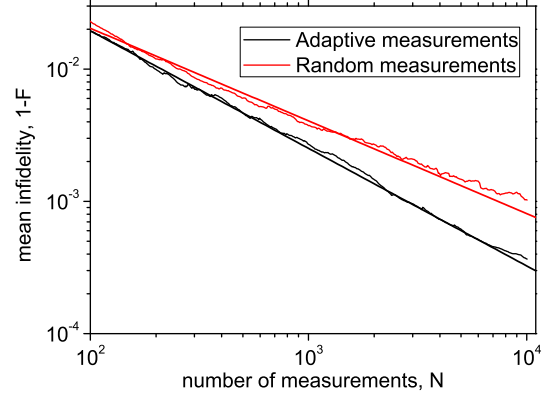


Figure 5: Experimental results: mean infidelity $1 - \hat{F}(\rho, \hat{\rho})$ with current Bayesian mean estimate $\hat{\rho}$ for non-adaptive measurements (black line) and adaptive measurements (red line). Data points are averages over 6 experimental runs. Solid straight lines are best power law fits.

we intentionally *did not* average over many realizations at each step of the algorithm, data points in Fig. 5 are average over several *full runs* of the experiment. The convergence rate behaves regularly from run to run. Fit-

ting the data averaged over 6 realizations with power law of the form $1 - \hat{F} \propto N^a$, we obtained $a = -0.700 \pm 0.005$ for non-adaptive protocol, and $a = -0.889 \pm 0.003$ for adaptive one.

In a real world application of tomography the 'true' state is unknown, and the Bayesian estimate given above is the only figure of merit at hand. However, in our experiments, we can estimate the prepared state independently. To do this we performed a very large number of Stokes parameters measurements. Let us denote the 'true' state, estimated this way, $\bar{\rho}$, we may now analyze the protocol performance using the mean infidelity with 'true' state $1 - \hat{F}(\rho, \bar{\rho}) = \mathbb{E}_{p(\rho|\mathcal{D}_n)} [1 - F(\rho, \bar{\rho})]$. The scaling with N of this quantity is depicted in Fig. 6. Power law fits in this case give $a = -0.502 \pm 0.001$ and $a = -0.902 \pm 0.008$ for the random and adaptive strategies, respectively. Error bars here are from the fit of an average curve, and within errors experimentally obtained scaling laws agree with simulations.

Our model does not take into account systematic errors caused, for example, by inaccuracies in waveplates rotation. However for the reached values of infidelities of on the order of $10^{-4} - 10^{-3}$ we did not observe any deviations from expected behavior and were not able to identify the influence of systematic errors. Further investigation with much larger statistics is required to address this issue.

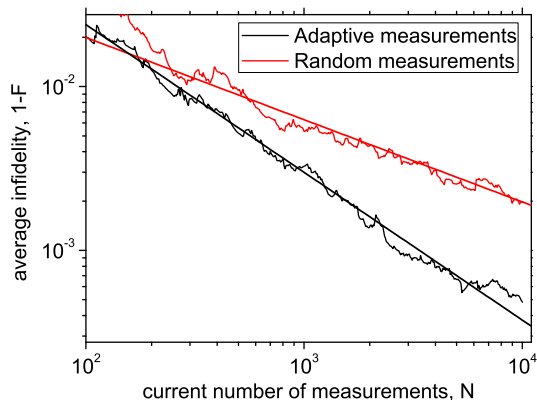


Figure 6: Experimental results: mean infidelity $1 - \hat{F}(\rho, \bar{\rho})$ with "true" state $\bar{\rho}$ for non-adaptive measurements (black line) and adaptive measurements (red line). Data points are averages over 6 experimental runs. Solid straight lines are best power law fits.

Conclusion. Our experimental results clearly demonstrate the advantages of adaptive strategies in quantum state tomography. We have adapted Bayesian methods of state estimation, because Bayesian methods maintain confidence levels, and error bars with their estimates, they are a very natural tool for the task of adaptive experiment design. Besides the aforementioned favorable

properties, the Bayesian approach is convenient from a purely practical point of view. It does not require any additional precomputation, and since posterior updates may be easily carried out after a single detection event, we expect that this approach will be particularly useful in the case of extremely weak signals. The N^{-1} scaling of infidelity in the adaptive case is the theoretical limit for any tomographic protocol, and further improvement may only affect pre-factors in this power law. Simulation results show that our strategy of choosing adaptively between general measurements outperforms any non-adaptive protocol, and although only results for completely random measurements and fixed bases are provided here, experimental work showing worse performance of more sophisticated non-adaptive strategies is underway. Finally, let us note that using an attenuated laser source is absolutely equivalent to a true single-photon source for the purposes of this particular single qubit experiment. Generalization of the developed adaptive protocol for two-qubit polarization states, and higher-dimensional systems (like spatial modes of the biphoton field) will be reported elsewhere.

After this paper was completed we have become aware of a highly relevant work [14] taking a different approach to adaptive state estimation and achieving similar performance.

This work was supported in part by Federal Program of the Russian Ministry of Education and Science (grant 8393) and state contract 911.519.11.4009, European Union Seventh Framework Programme under grant agreement n0 308803 (project BRISQ2), ERA.Net RUS Project NANOQUINT, grant for the NATO project EAP.SFPP 984397 "Secure Communication Using Quantum Information Systems" and by RFBR grants 12-02-31041 and 12-02-31792. S. S. Straupe is grateful to the "Dynasty" foundation for financial support. N. Houlshy is grateful for support from The Google European Doctoral Fellowship programme.

* Electronic address: straups@yandex.ru

- [1] M.Paris and J.Rehacek, *Quantum state estimation*, Lect. Notes Phys. **649** (Springer, 2004).
- [2] S.L. Braunstein, C.M. Caves, Phys.Rev.Lett, **72** 3439 (1994).
- [3] I.Bengtsson and K.Życzkowski, *Geometry of quantum states*, (Cambridge University Press, 2006).
- [4] Z.Hradil, Phys. Rev. A, **55** 1561 (1997).
- [5] J.Rehacek, B.-G.Englert, and D.Kaszlikowski, Phys.Rev.A **70**, 052321 (2004).
- [6] M.D. de Burgh, N.K.Langford, A.C.Doherty and A.Gilchrist, Phys.Rev.A **75**, 052122 (2008).
- [7] M.Genovese, G.Brida, E.V.Moreva, A.P.Shurupov, A.A.Soloviev, K.S.Kravtsov, Yu.Bogdanov, I.D.Bukeev, S.P.Kulik, Phys. Rev. A, **84**, 042108 (2011).
- [8] D. G. Fischer, S. H. Kienle, and M. Freyberger, Phys.

- Rev. A **61**, 032306 (2000).
- [9] T. Hannemann, D. Reiss, C. Balzer, W. Neuhauser, P. E. Toschek, and C. Wunderlich, Phys. Rev. A **65**, 05030 (2002).
 - [10] F. Huszár and N. M. T. Houlsby, Phys.Rev.A **85**, 052120 (2012).
 - [11] R.Okamoto, M.Iefuji, S.Oyama, K.Yamagata, H.Imai, A.Fujiwara, and S.Takeuchi, Phys.Rev.Lett. **109**, 130404 (2012).
 - [12] Robin Blume-Kohout, New Journal of Physics **12**, 043034 (2010).
 - [13] J. Dittmann, J.Geom.Phys. **31**, 16 (1999).
 - [14] D.H.Mahler, L.A.Rozema, A.Darabi, C.Ferrie, R.Blume-Kohout, and A.M. Steinberg, arXiv:1303.0436v1 [quant-ph] (2013).
 - [15] Lvovsky, Hansen, Aichele, Benson, Mlynek, and Schiller, Physical review letters **87** (2001).
 - [16] R. Blume-Kohout, New Journal of Physics **12**, (2010).
 - [17] K. M. Audenaert, S. Scheel, New Journal of Physics, **11**, 023028 (2009).
 - [18] A. Doucet, N. De Freitas, N. Gordon, et al., *Sequential Monte Carlo methods in practice*, Springer New York, (2001).
 - [19] S. Panzeri, R. Petersen, Journal of neurophysiology **98**, 1064 (2007).
 - [20] Hannemann, Reiss, Balzer, Neuhauser, Toschek, and Wunderlich, Physical Review A **65**, 050303, (2002).
 - [21] W. Wootters, B. Fields, Ann. Phys. **191**, 363(1989)
 - [22] Sometimes, eg. in [7] this quantity is called *fidelity loss*.